Applied Technologies, Inc.

FINAL REPORT

ON A PC-BASED DISK ARRAY SUBSYSTEM

US NAVAL AIR WARFARE CENTER

POINT MUGU, CALIFORNIA 93042

CONTRACT NUMBER N00244-95-C-0079

DATES OF PERFORMANCE:
FEBRUARY 2, 1995 to AUGUST 31, 1995

THIS DOCUMENT IS APPROVED FOR PUBLIC RELEASE
DISTRIBUTION IS UNLIMITED

PREPARED BY:

APPLIED TECHNOLOGIES, INC.
6395 GUNPARK DRIVE, UNIT E
BOULDER,COLORADO 80301

DTIC QUALITY INSPECTED 1

20010125 100

TABLE OF CONTENTS

Approved for Public Release; Distribution is Unlimited

1. INTRODUCTION

Upon receiving the US Navy SBIR contract award, Applied Technologies, Inc. contacted the technical personnel at the Naval Air Warfare Center and arranged for a meeting to fully understand the requirements of the SBIR study. Following the technical discussions, Applied Technologies ordered two Pentium computer to begin testing interface boards to determine the data rate throughput between the two computers. It was determined that the Fast Ethernet interface boards throughput limit is just over 2 MBytes per second.

The remaining task was to determine the data recording and replay speeds. Applied Technologies provided five, 1 GigaByte SCSI drives for further testing of a RAID 3 configuration. Tests with three RAID system configurations indicated a limitation of 1 MBbyte record and 1 MBbyte playback. It was determined that this limitation was partially due to a shortcoming of the Microsoft NT operating system. By reformatting the controller to accommodate this shortcoming, a final test at a vendor's facility indicated a record rate of greater than 5 MBytes per second and a playback rate of greater than 2 MBytes per second.

It would appear that the ability to record and playback data at the rate required by the Naval Air Warfare Center is feasible with the NT operating system. The problem outstanding is the input interface limitation.

2. TECHNICAL OBJECTIVES

The standard technique for recording telemetry data in range operations in the past has been instrumentation tape recording. This has provided the capability for recording data at high data rates (>2 Mbytes/sec) from multiple sources for extended periods of time (hours, if needed). While very convenient and cost-effective for recording, this technique suffers significant drawbacks in subsequent processing and playback of the data. Due to the serial nature of the recording medium, no capability exists for near real-time playback of data unless a separate, dedicated processing system is used. Turnaround of data packages recovered from tape recording can take on the order of days which is often unacceptable.

The technical objective of this study was to evaluate an alternative to tape recording which potentially could resolve the playback difficulties. A primary constraint on this alternative was that it be low cost. The cost constraint eliminated the possibility of using random access memory (RAM) for the data acquisition platform. The other possibility was to use disk drives as the storage medium.

Significant advances have occurred within the computer industry in the last few years in disk drive technology for data storage. Media transfer rates within disk drives have climbed as rotation rates for media now exceed 7200 RPM. The cost of disk drives have plummeted, with prices nearing $0.25 per Mbytes of storage. Such drives are now available in the 3.5 inch form factor with capacities of 4 Gbytes.

Most important for this opportunity addressed was a new method of combining independent drives to operate functionally as a single large disk drive. This technology is referred to as RAID, an acronym for Redundant Array of Independent Disks. The primary technical objective of this study was to evaluate RAID systems connected to a high-end Pentium PC to determine if these systems could sustain the continuous data acquisition rates required by range operations.

While RAID technology is ideally suited to meet the storage requirements of this opportunity, two more issues must also be addressed to complete the system. The first issue requires that we ensure that the host platform has the required data bandwidth to process data at the rates required. Secondly, we must provide a user-friendly means of accessing the data after recording to ensure that data products can be provided quickly and accurately.

Approved for Public Release; Distribution is Unlimited

In the initial proposal we also indicated that we would need to work with the customer to determine the most effective means of getting data into the system prior to storage on a RAID system. Based on discussions with range personnel, we additionally examined Fast Ethernet performance on a Pentium PC.

Based on the results of the technical evaluation of the platform described above, the second technical objective for Phase I of this opportunity was to design and submit a proposed solution specifying the technology to be utilized. Sufficient testing would be done with leased hardware to demonstrate the feasibility of any approach taken. Since there is a lot of variability in how manufacturers' specifications are developed, we intended to test with real hardware any critical design parameters. This design is provided in this report. Included with the design we provide cost estimates for the final product. This includes not only the base product, but possible extensions and upgrades. Spare parts and support costs are also indicated.

3. WORK EXECUTION

3.1 WORK ORIGINALLY PLANNED

The first part of the work plan was to visit with Sea Range personnel to develop a clear understanding of their requirements. We considered it important to gain a clear understanding of the requirements, but did not intend to get overly involved in developing a rigid written requirements document.

From the overall system requirements, we were to develop a break down of those requirements to specific subsystems. We expected the subsystems to consist of the following:

1. I/O interface for the input of telemetry data.
2. Dedicated processor board with PCI bus and associated real-time kernel.
3. Disk Array controller board. PCI bus to SCSI channels.
4. User Interface to platform.
5. Data Display Device or I/O interface to existing Sea Range displays and/or data package generators
6. The basic PC platform--Chassis, Power Supply, Cooling, Cabling, etc.
7. Disk Array packaging platform--Chassis, Power Supply, Cooling, Cabling, etc.
8. Disk Drives to populate the array.
9. Software Requirements Documentation for specific code to be generated for this platform.

Figure *3-1* shows a diagram of the elements of the originally proposed system. Data was to come in from telemetry link through an I/O card into the PC memory. In conversation with the customer, the means of data input we were to evaluate became instead a Fast Ethernet link to existing Range hardware. A processor card dedicated to the write operation was envisioned to control the movement of data from the PC memory to the RAID 3 disk array controller. This became the central Pentium processor in a commercial PC. Windows NT became the operation system of choice. A disk array processor for RAID 3 transfers directly from the internal PCI bus has not become available at this time, so we evaluated standard SCSI Host Bus Adapters (HBA) which plugged into the PC and then communicated with a dedicated RAID controller over the SCSI bus.

Approved for Public Release; Distribution is Unlimited

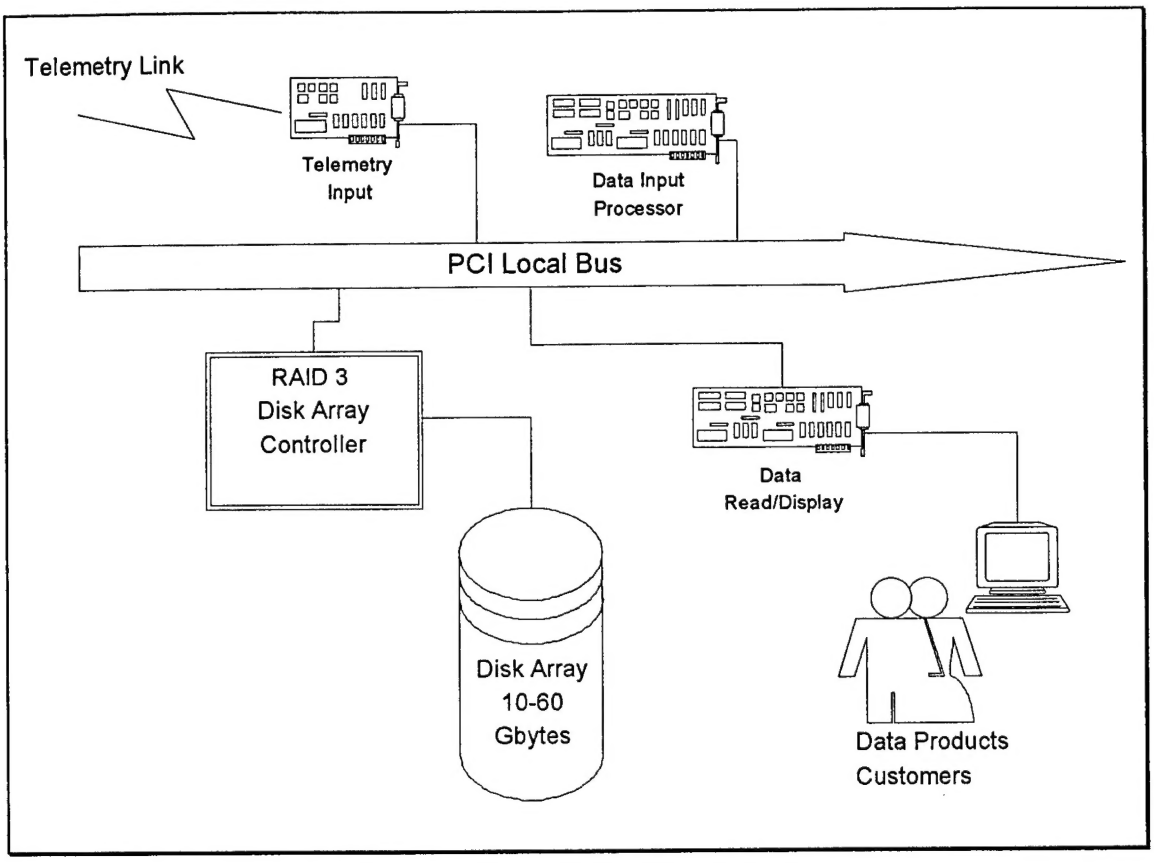

Figure 3-1 Diagram of Originally Proposed System

In the original proposal we also proposed a second processor to handle the user interface and data read requests. Due to difficulties in evaluating the Fast Ethernet board and a reduction in the desired scope of display capabilities of this data acquisition platform, the second processor possibility was not explored.

The RAID 3 disk array controller was critical to the success of the system. We expected to spend considerable time researching options for this subsystem. RAID Level 3 controllers are available from a number of vendors with highly varying cost and performance factors. We evaluated three primary systems.

We expected that the disk drives to populate the array would require careful analysis and possible testing to ensure that the disk drives chosen will work in this system. We anticipated using drives with a SCSI interface. This turned out to be less critical than anticipated as the RAID controller vendors had already spent considerable time qualifying drives for their systems.

Critical to meeting system requirements is the response of the system to an error condition. While we would anticipate that telemetry data would still be recorded to instrumentation tape such that no data would be lost, we need to ensure that the system does not go down on a single error and remain down for the duration of the mission. Areas that will required examination are the array controller card, the drives themselves, and software developed by Applied Technologies, Inc.

We expected in the original proposal, and still do, that a large percentage of the work in Phase II would be software development for the operation of the system. As will become clear in this report, the majority of problems encountered were with software, notably the Windows NT Operating System and device drivers.

Approved for Public Release; Distribution is Unlimited

The next step after having a clear and complete understanding of system and subsystem requirements was to survey the marketplace and determine the availability and suitability of commercially available components from which to build the system. Applied Technologies, Inc. did not intend to consider designing specific board-level PC components for this opportunity unless there proved to be no acceptable components available. We have determined that it is possible to build this system with commercially available hardware, but have found problems with standard software. The primary activity still anticipated for Phase II is system integration and development of software to support integrated system operation.

Having determined the subsystem components suitable to design this system, we obtained initial test data using evaluation units to determine if the components would meet design requirements. This test system was not a complete system, rather a smaller platform to test critical parameters. Test results are included in this report.

3.2 ADDITIONAL WORK

As a result of the coordination meeting with the Naval Warfare Center technical personnel, we determined that some effort needed to be expended in investigating the data input characteristics of the recording system. The Naval technical personnel requested that the input medium desired was a Fast Ethernet system, and the operating system to be Windows NT.

To explore the data input requirement it was necessary to purchase two Fast Ethernet cards and two high performance computing systems with PCI bus capabilities.

3.3 RESULTS OF WORK

3.3.1 Pentium/PCI PC Hardware Performance

The computer systems used to perform the testing were Pentium 100 Mhz, which should have been more than adequate to perform the testing required. The computer had sufficient PCI slots to accommodate the Ethernet cards and other SCSI adapter cards. The PCI bus was tested and shown to be able to handle at least 22 MBytes/second and was therefore adequate for the systems testing.

3.3.2 Ethernet Card Hardware Performance

In order to investigate the Navy request for a Fast Ethernet medium, interface cards were purchased from Cogent Data Technologies. These interface cards were advertised as capable of transmitting and receiving 8 megabytes/second.

We determined that the most efficient way to determine data throughput was to use two computers and transfer data from one computer to the other through a Fast Ethernet medium. Two computers were set up in a client/server configuration with software written to transfer a 1 MBbyte buffer from one machine to the other. This was essentially a memory to memory transfer through the two Fast Ethernet interface cards.

Several protocols were used to in the tests: IPX, NetBEUI, and TCP/IP. Each of the protocols produced similar results, however none exceeded 2.2 MBytes/second. The NetBEUI produced the best results, but only slightly better than the others. The results during the testing phase did not provide the same results that were advertised by Cogent, in fact they were much lower than expected. The software test program was sent to Cogent who ran the program and arrived at the same results that were seen at Applied Technologies. Conversations with Cogent did not reveal any reason for this discrepancy.

Approved for Public Release; Distribution is Unlimited

Technicians at Cogent experimented with different Window NT drivers and were not successful in improving performance. The Cogent technicians talked with some Microsoft technicians and discovered that there are some throughput limitations. However, neither company would accept responsibility for the limitation. Using a program called NetBench Cogent was able to show data throughput 5 megabytes per second.

3.3.3 SCSI HBA Performance

A SCSI host adapter card was purchased from Adaptec to perform the Disk Array write and read tests. This card was used in the majority of the testing with different vendor's controller systems. Later in the testing, a BusLogic SCSI card was used and provided slightly, but not significant, better performance.

3.3.4 Disk Array Testing Results

At the beginning of the Phase I study, we determined that there were three notable manufacturers of Disk Array Systems. The companies that were chosen to be evaluated were:

- CMD
- RAIDTEC
- CIPRICO

These manufacturers are known to produce a good products and are aggressively pursuing the Disk Array storage business. CMD and Raidtec agreed to supply a system for testing, either at our facility or at one of their distributor facilities. The CMD and Raidtec disk array systems are similar in price, while the Ciprico is more expensive.

Testing Methodology

A disk array system was configured as a RAID 3, with five, 1 Gigabyte Seagate Hawk drives. A software test program was written that transferred a 40 MBbyte file to a disk array system by incrementing the block size 4 kilobytes/second at a time.

The software test program was written by Applied Technologies and was therefore available for vendor testing at their facilities when we could not participate in the testing.

A more comprehensive test was devised that tested a composite, write 5/read 2 MBytes/second data transfer. Data were gathered for each of the tests on the various array systems and were plotted.

Approved for Public Release; Distribution is Unlimited

CMD Performance

A CMD model CRD 5000 Disk Array Controller was not available for testing at the Applied Technologies facilities, however a CMD distributor, Data Storage Marketing, had a system available and were willing to accommodate our engineer in running tests.

The results of the Data Storage Marketing testing indicated results of approximately 3 - 3.5 megabytes/second write and 4 - 4.5 MBytes/second read.. A second test was run where the ratio of write to read was 5/2 with varying block sizes. The results of these tests are shown in the following three figures, (Figure 3-2, Figure 3-3, and Figure 3-4).

The software test program was sent to CMD in an effort to determine if we were doing something wrong. They ran the software on a CRD 5000 system in their facility with the same results. We tried unsuccessfully to get them to run the software on a prototype/engineering model of their new CRD 5500. The CMD technical support personnel were unable to get the engineering department to release the prototype for testing. At this time we do not know how much better the new controller over that of the production CRD.

Approved for Public Release; Distribution is Unlimited

Approved for Public Release; Distribution is Unlimited

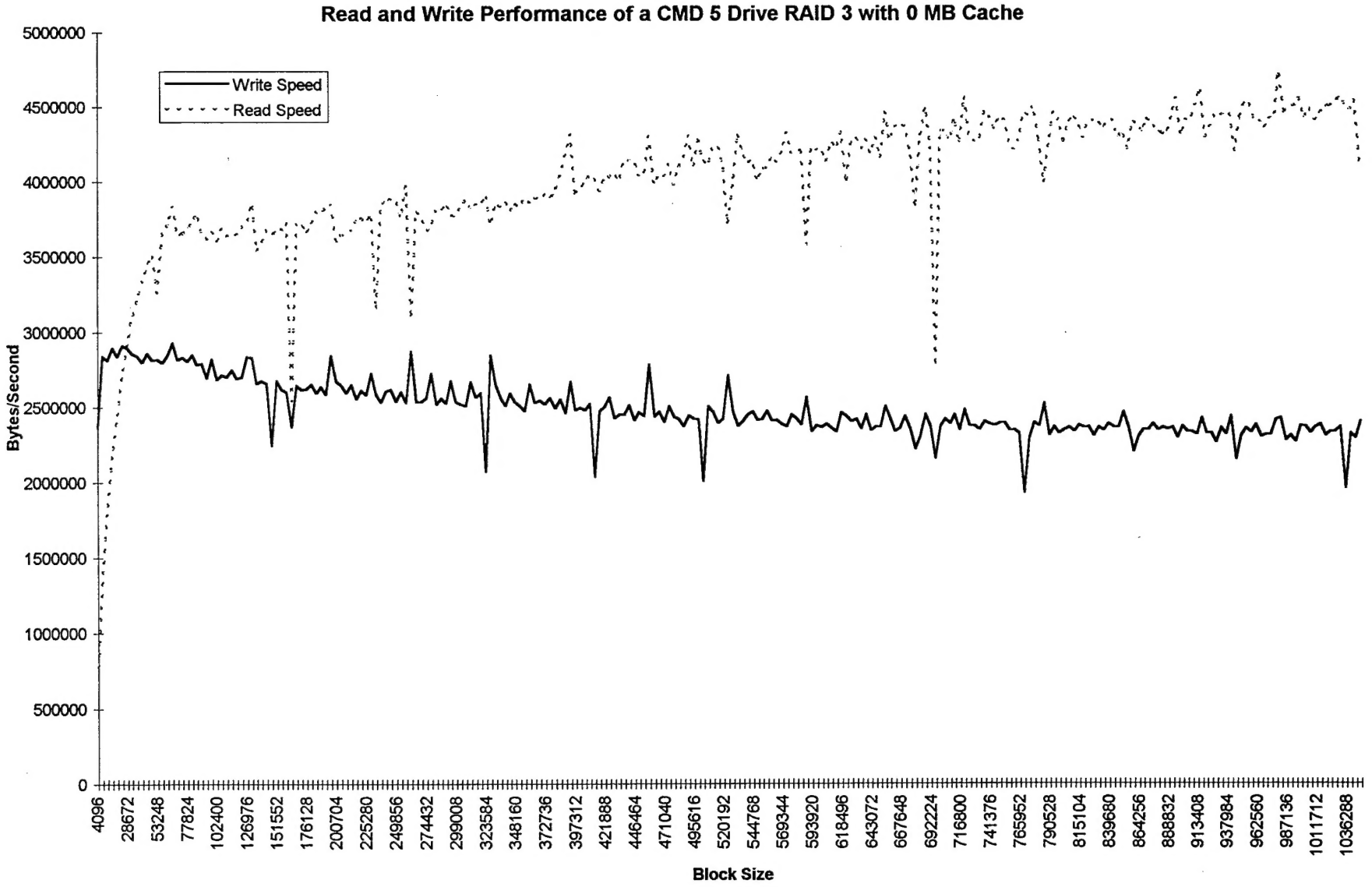

Figure 3-2

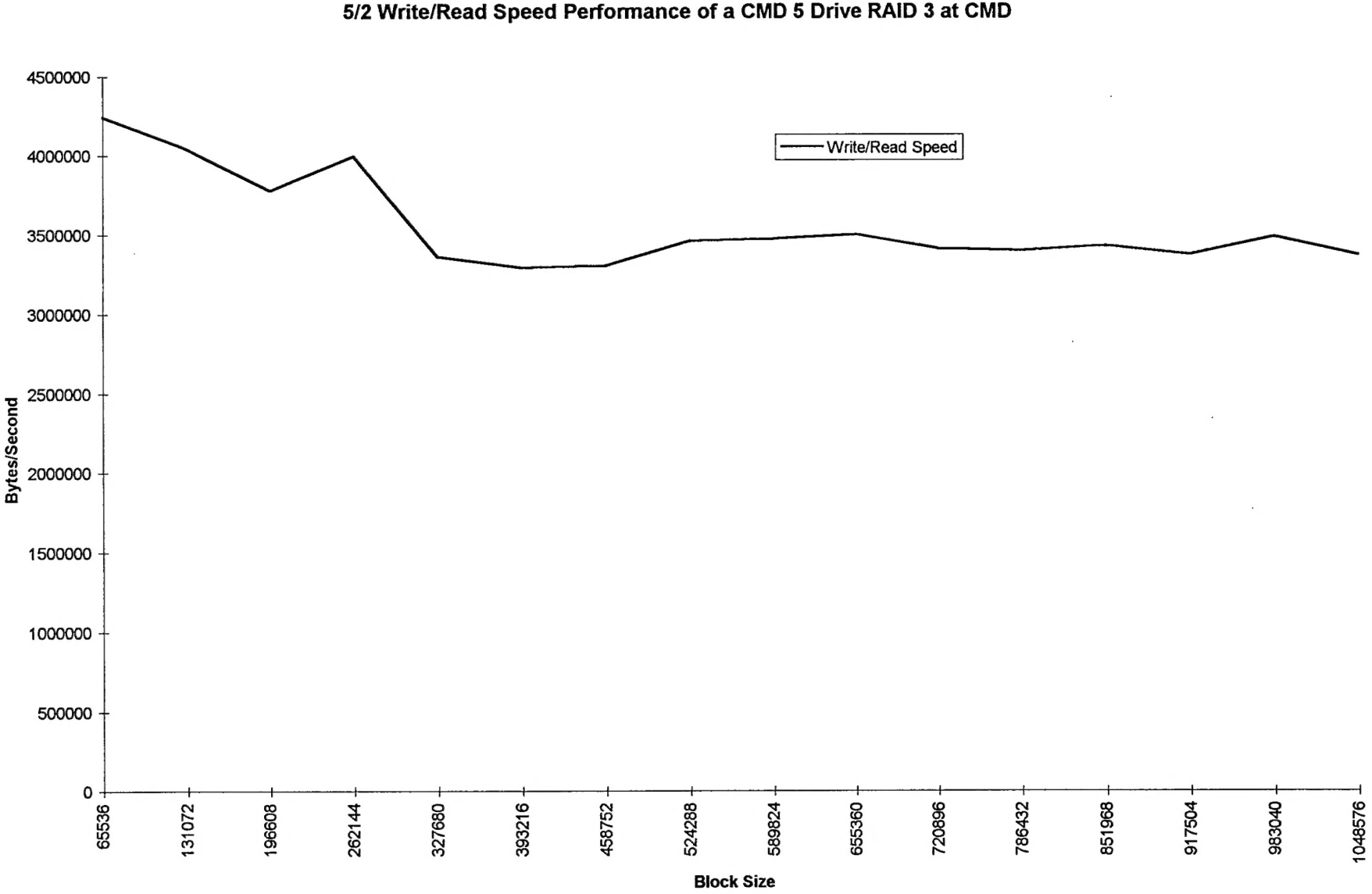

Figure 3-3

Approved for Public Release; Distribution is Unlimited

Approved for Public Release; Distribution is Unlimited

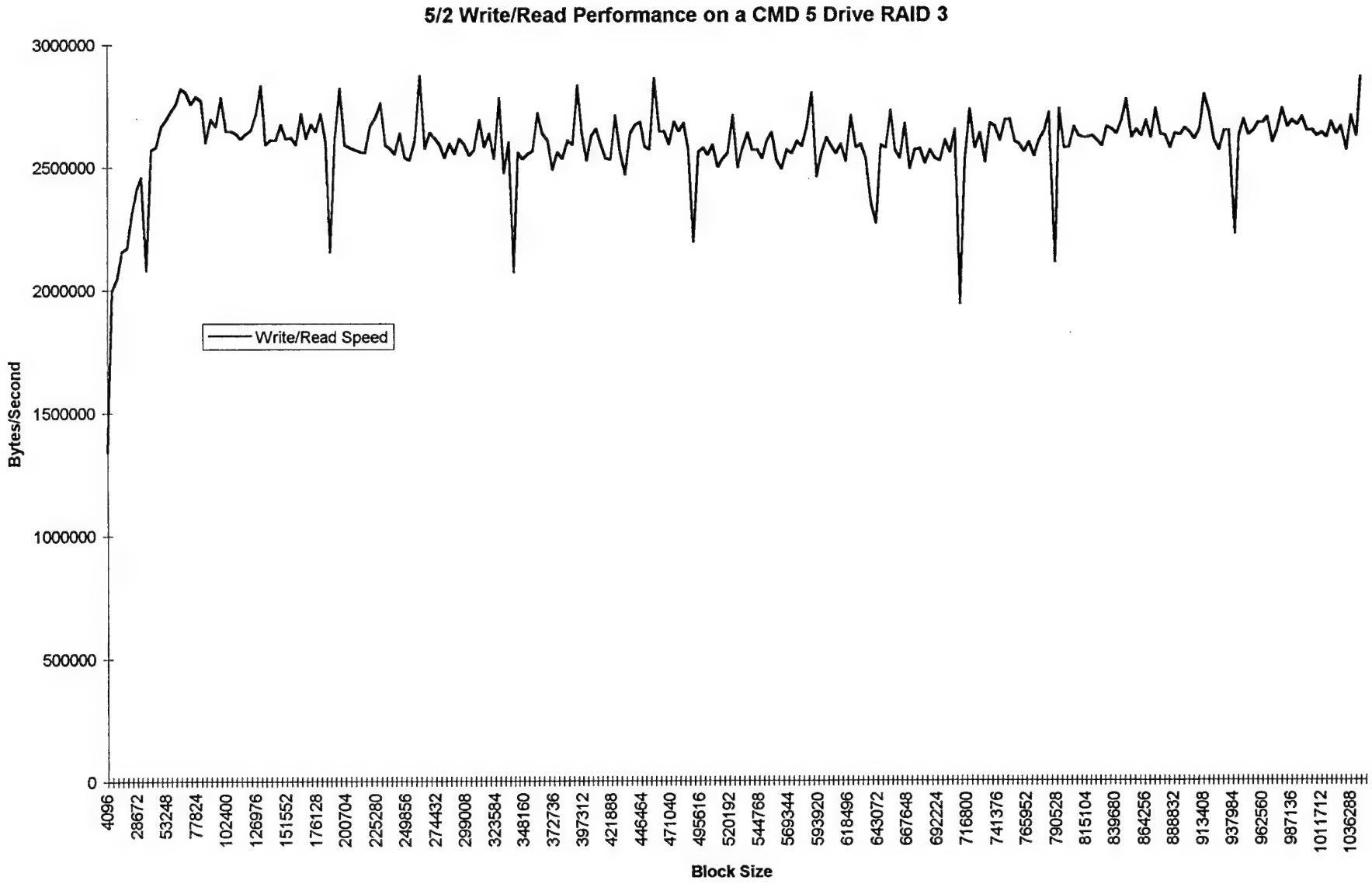

Figure 3-4

RaidTec Performance

Raidtec offered to let us evaluate one of their disk array controllers for a period of one week. We used our five, 1 Gbyte Seagate Hawk disks in their array controller. The same set of tests were run on the Raidtec disk array system. Results were not as good as that with the CMD. Data were gathered and plotted The results of the testing are shown in Figures 3-5, 3-6, and 3-7.

The software test program was sent to Raidtec in an effort to get them to run it on their in-house systems. All efforts to get them to provide data were unsuccessful.

Approved for Public Release; Distribution is Unlimited

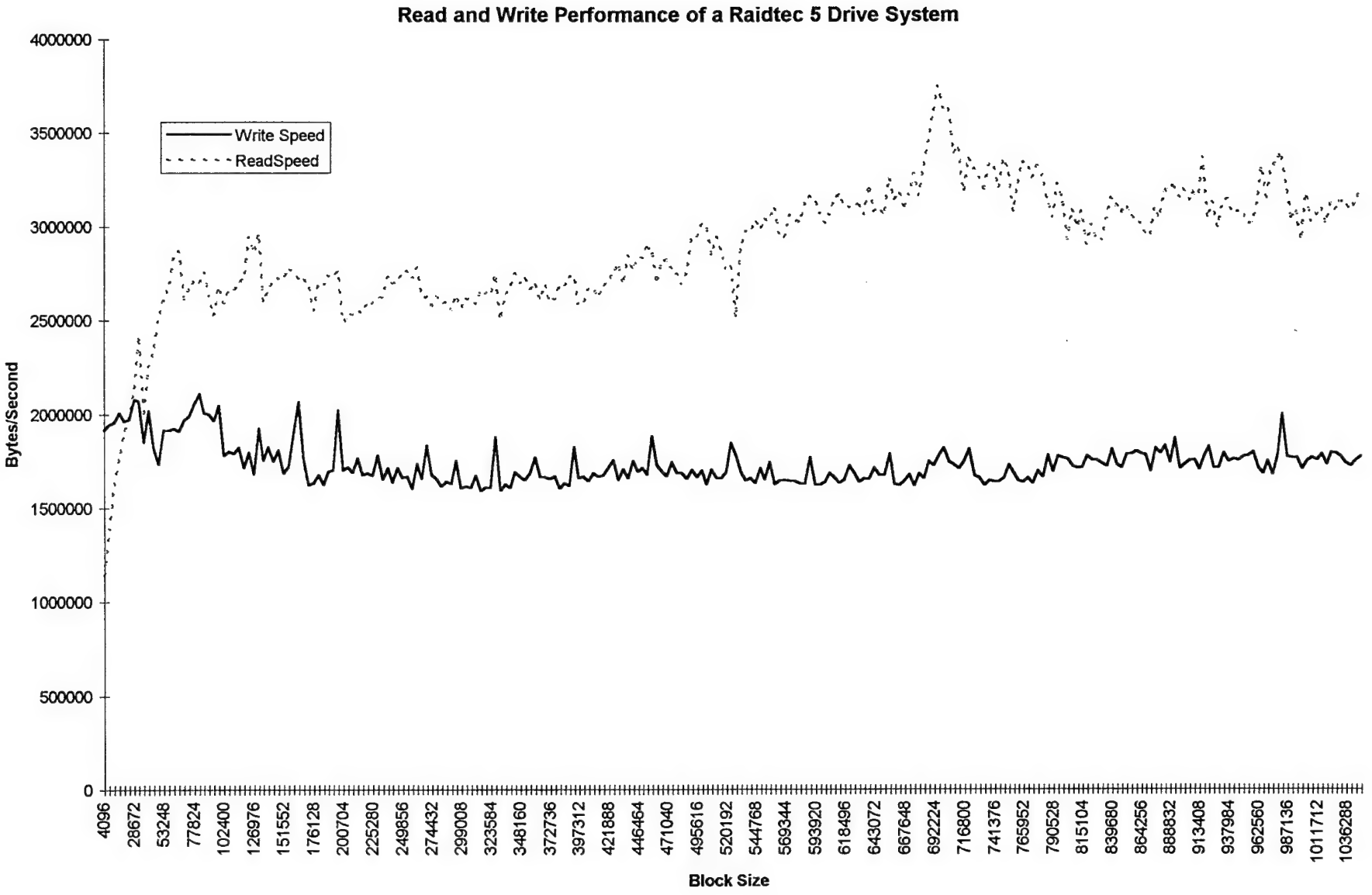

Figure 3-5

Approved for Public Release; Distribution is Unlimited

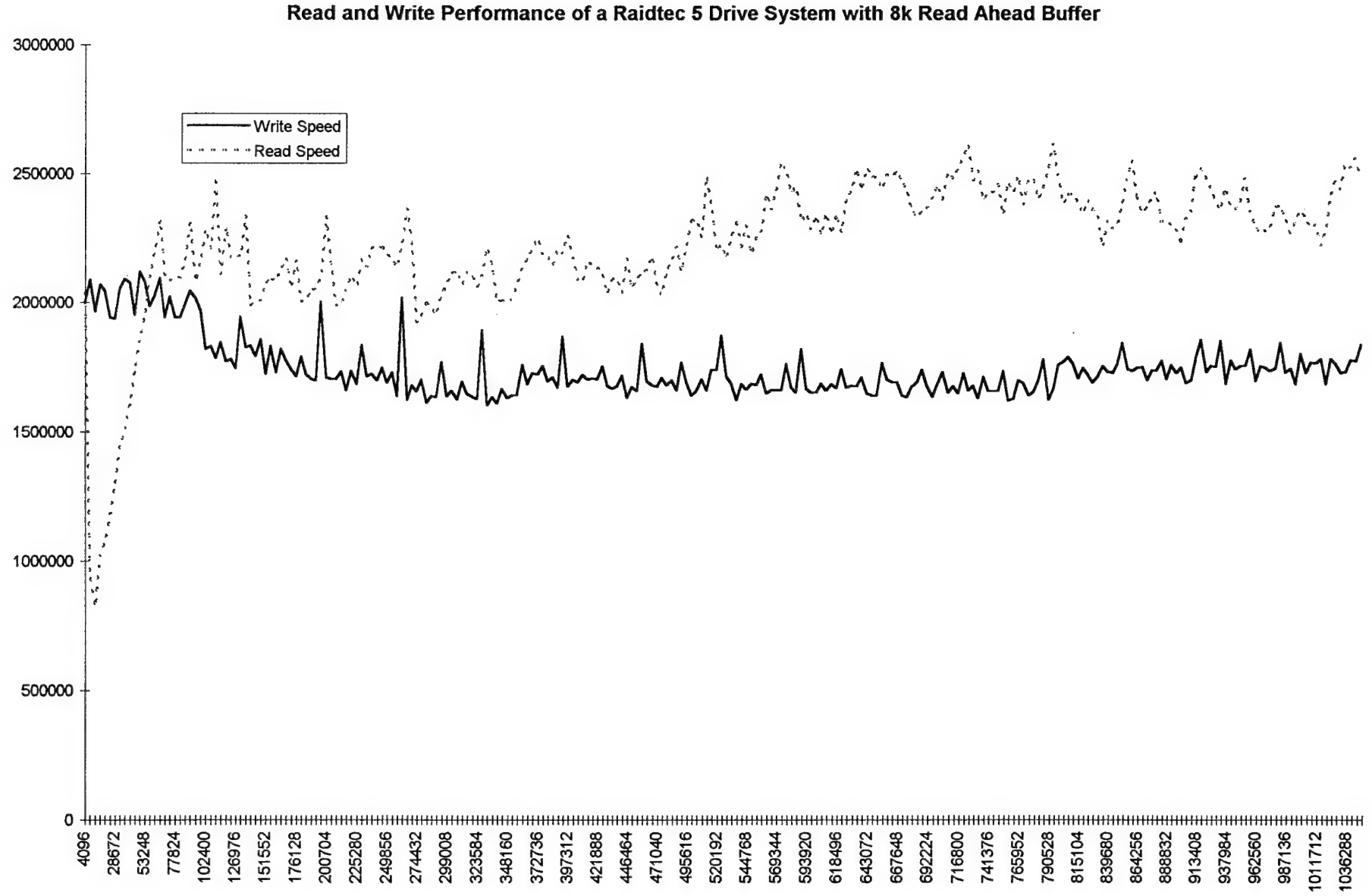

Figure 3-6

Approved for Public Release; Distribution is Unlimited

Approved for Public Release; Distribution is Unlimited

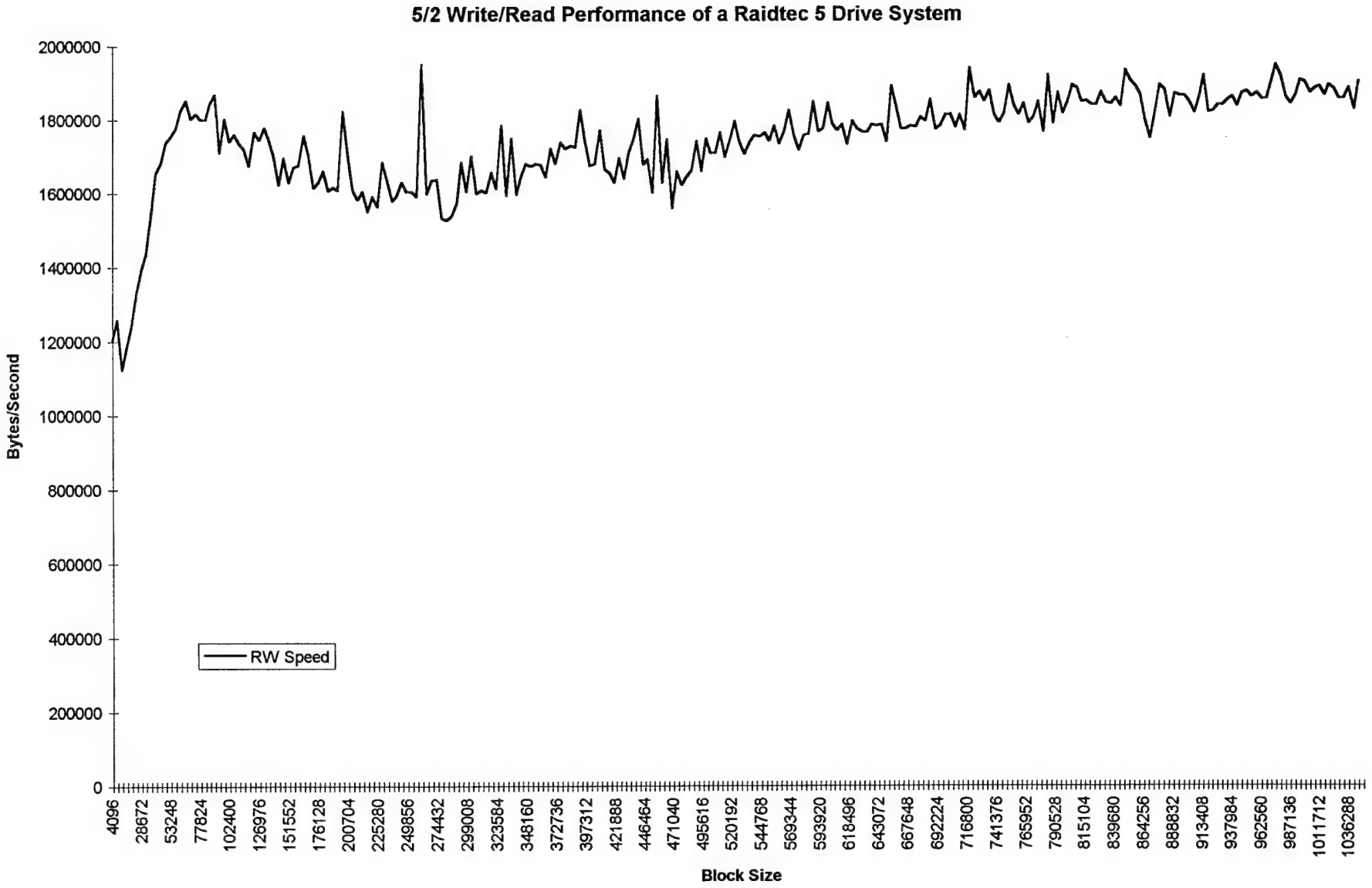

Figure 3-7

Ciprico Performance

Ciprico would not consent to send a disk array system for evaluation, however their technical support personnel consented to run the software test program on their own in-house system. Their system consisted of five, 2 GByte Barracuda drives. Initial results were very similar to the other disk array manufacturers.

By attaching a SCSI Analyzer to the disk array controller, the Ciprico technical personnel discovered that Windows NT has a peculiarity in its architecture that only allows 64 Kbytes to be transferred at a time. By reformatting their controller to accommodate the 64 Kbyte data block size, they were able to show that the reformatted system could write at approximately 9 Mbytes/second and read between 14 - 15 Mbytes/second as shown in Figure 3-8.

We were impressed with the Ciprico initiative to investigate the cause of the problem. Apparently the Windows NT will not at this time allow larger block size transfer than 64 KBytes.

Approved for Public Release; Distribution is Unlimited

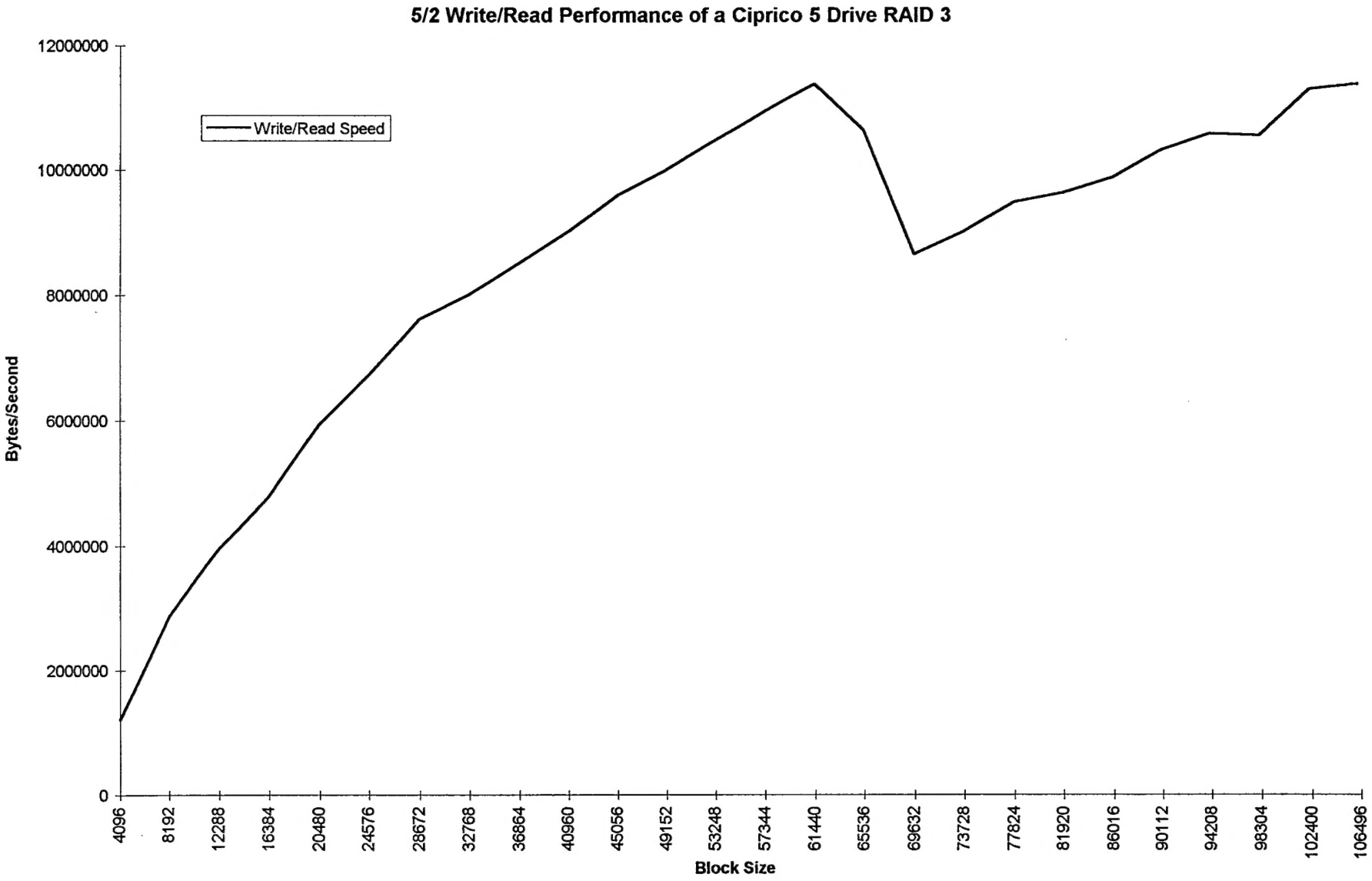

Figure 3-8

Approved for Public Release; Distribution is Unlimited

3.3.5 Disk Array Products Status

There will be a Comdex show in November where there will be a number of vendors announcing second generation RAID controllers (not just CMD). These new controllers could provide the performance of the Ciprico box at lower cost. We would intend to attend this show as part of the Phase II effort.

3.3.6 Windows NT Overall Performance

With the limitation of the Windows NT software, this package may not be the ideal software system to use in this application. Earlier discussions with Microsoft have shown a unwillingness to change or modify any of their software. As a part of Phase II an investigation of other software packages should be initiated. We have identified two possible low-cost operating systems: Linux and FreeBSD. These are only two examples of systems that we have in-house for evaluation.

3.3.7 Ethernet Software Driver Limitations

The poor results of the Cogent Technologies, Fast Ethernet card was certainly a surprise to the Cogent personnel. It is conceivable that the poor performance of the Fast Ethernet cards may be due in some way to the limitation the Windows NT.

It may be constructive to investigate interface methods other than Ethernet. This should be part of the Phase II effort.

3.3.8 SCSI Driver Limitations

The SCSI Fast and Wide cards that are on the market should be more than capable of dealing with the sustained data write and read rate. Again the Windows NT limitation may be the bottleneck to providing a clean data transfer rate.

3.4 SUMMARY OF RESULTS

The results of the SBIR study to determine the capability of a PC-Based disk array system to record a sustained 5 Mbytes/second and read near real-time 2 Mbytes/second are positive, given a method of interfacing the data to the high performance PC. The PC system used in the study is an off-the-shelf Pentium, operating at 100 Mhz, using Windows NT as the operating system. Data have been gathered and plotted indicating tests and performance results of three manufacturer's disk array systems.

Only one of the systems actually demonstrated that they could write a sustained data rate of greater than 5 Mbytes/second and read the data from the disk array at a rate greater than 2 Mbytes/second. This system, CIPRICO, consisted of a RAID 3 disk array configured as 4 plus 1, with 2 Gbyte Barracuda drives.

A CIPRICO disk array system can be configured that would provide approximately 17 Gbytes of data storage, consisting of 4 plus 1, 4 Gbyte Barracuda drives, with redundant power supplies. This system would cost approximately $31K. A dual-Pentium computer operating at 133 Mhz would cost approximately $5.2K. For a total hardware cost of less than $40K a system could be provided to the US Naval Air Warfare Center.

An alternative system capable of storing 34 Gbytes of data would cost approximately $53K. This system would consist of 8 plus 1, 4 Gbyte Barracuda drives. The total data storage requirement of the US Naval Air Warfare Center will determine the ultimate disk array configuration. Total data storage is a function or input data rate and data record time as shown in Figure 3-9.

Our recommendation is to continue the effort in Phase II and provide a deliverable Disk Array System to the US Naval Air Warfare Center. With the knowledge gained during the Phase I study effort, a system that meets the requirement can be designed and delivered.

Approved for Public Release; Distribution is Unlimited

Approximate Storage Duration in Minutes

Input Data Rate Mbytes/s	Capacity in Gigabytes							
	5	10	15	20	25	30	35	40
1	83	167	250	333	417	500	583	667
2	42	83	125	167	208	250	292	333
3	28	56	83	111	139	167	194	222
4	21	42	63	83	104	125	146	167
5	17	33	50	67	83	100	117	133
6	14	28	42	56	69	83	97	111

Figure 3-9

Approved for Public Release; Distribution is Unlimited